

언론정보연구  
48권 1호, 2011년, 188~228  
<http://icr.snu.ac.kr/jcr>  
서울대학교 언론정보연구소



## 대한매일신보 국문논설 번역문체 판별의 어휘적 준거에 대한 탐색적 연구 ‘형태주석 말뭉치’의 어휘를 중심으로

윤상길

인천대학교 인천학연구원 박사후연구원 cyrus92@naver.com

김영희

서울대학교 언론정보연구소 책임연구원 properk@hanmail.net

최운호

성공회대학교 강사 woonho\_choi@hotmail.com

본 연구는 『대한매일신보』 국문논설 형태주석 말뭉치를 활용한 사례연구로서, 말뭉치에 나타난 어휘들의 계량적 특성에 주목하여, 논설의 번역문체를 판별하는 어휘적 준거를 탐색하는 데 그 목적이 있다. 보다 구체적으로는, 기존 연구에서 저자판별을 위한 문체소적 자질로서 검증되었던 비주제 특정 형태소(non-topic-specific morpheme) 중의 하나였지만 검증이 되지 않았던, 형용사와 부사에 초점을 맞추어, 각 어휘들의 빈도가 시기별로 어떠한 식으로 변모해 가는지를 전체적으로 살펴본다.

연구결과, 양기탁에 의해 영문판논설에 대한 국문논설번역이 이뤄지던 시기의 부사 어휘와 이후 시기의 부사 어휘가 유의미하게 차이가 있다는 점을 고빈도 부사 어휘에 대한 분석으로 통해 밝혀냈는데, 이를 통해 (대한매일신보 국문논설에만 국한시켜 적용시킨다는 전제 하에) 기능어 어휘 중에서 부사 어휘가 문체적 특성을 나타내는 문체소적 자질이 될 수 있음을 확인하였다.

**KEYWORDS** 어휘적 준거 • 문체소적 자질 • 저자판별 • 형태소 • 말뭉치

## 1. 머리말

이 글은 근대이행기 신문 중의 하나이자 1900년대 중반 이후에 발행된 대표적인 신문인 『대한매일신보』 국문논설을 대상으로 구축된 형태주석 말뭉치(annotated corpus)를 활용한 사례연구로서, 말뭉치에 나타난 어휘들의 계량적 특성에 주목, 대한매일신보 국문논설의 번역문체를 판별하는 어휘적 준거를 탐색하는 데 그 목적이 있다. 보다 구체적으로는 형태주석 말뭉치의 시기별 어휘 특징에 대한 분석을 통해 문체를 판별하는 ‘형태소(形態素)’, 즉 ‘문체소(文體素)’를 도출함으로써, 향후 계량적 방법에 의한 ‘저자 판별연구(authorship)’의 방법론적 기초를 마련하고자 한다.

주지의 사실이지만, 대한매일신보는 한국사회에 근대적인 신문이 점차 하나의 사회제도로서 정착되고 (아직 보급부수가 많지는 않았지만) 신문독자들에 의해 신문구독 행위가 일상생활의 하나로 차츰 자리잡아가던 무렵인 1904년 7월 18일 창간되어 1910년 8월 우리나라가 일본에 의해 강제 병합되기까지 발행되었던 근대이행기의 중요 텍스트이다. 개화기로 불리는 19세기말 20세기 초의 시기는 한국사회의 근대이행기로서, 정치, 경제, 사회, 문화 등 모든 분야에서 근대적인 인식과 개념이 새롭게 형성되던 시기라고 할 수 있다. 외부적으로 선교사들과 이들이 반입하였던 서적들을 중심으로 서구열강 및 청국의 사상들이 유입됨으로써, 봉건적인 조선사회에서의 기존 인식과 개념이 변용되는 시기였다. 이런 측면에서 볼 때, 개화기에 등장하였던 근대의 텍스트들은 기존의 인식과 개념들이 새로운 근대적인 인식과 개념들과 만나 변화를 겪게 되는 하나의 담론적 장(場)이라 할 수 있다. 특히 신문은 각 분야의 공적인 문제가 공개적으로 논의되는 장으로서, 기존의 개념과 최신의 개념이 만나 새로운 제3의 개념이 사회적으로 형성되는 개념형성의 주요한 무대라 해도 과언은 아니며, 여러 분야의 근대적인 인식과 개념 형성의 추이를 분석하는데 있어서 당시의 신문들은 중요한 연구자료 임에 틀림없다.

외국에서는 근대이행기를 비롯한 특정시기의 텍스트들, 특히 신문

에 대한 계량적인 내용분석을 위해 컴퓨터를 활용하는 연구들이 매우 활발히 전개되어 왔다. 가령, 모스텔러와 월리스(Mosteller & Wallace, 1984)의 연구는 1787년 10월에서부터 1788년 8월까지 대부분 'The Independent Journal'에 무기명으로 연재되었던 신문논설을 대상으로, 계량적 '문체연구(stylometry research)'의 일환으로서 계량적인 내용분석 방법을 사용하여, 그동안 '연방제도'에 대한 사상형성과정에 있어서의 논쟁에 중지부를 찍는 데 기여하였다. 그 밖에도 근대이행기는 아니지만 서신이나 극본과 같은 텍스트를 대상으로 계량적인 내용분석을 한 연구들이 있다(Kenny, 1986 ; Holmes, 2001). 이처럼 외국에서는 계량적인 내용분석을 위해 컴퓨터를 활용하는 연구들이 매우 활발히 전개되어 왔지만, 국내에서 근대이행기 텍스트들을 대량적으로 내용분석하기 위한 기반연구들은 상대적으로 빈약한 실정이며, 이를 저자 판별 연구로 활용한 연구는 계량적인 분석방법을 사용하여 독립신문의 저자판별을 시도한 강남준 외의 연구(2010)를 제외하곤 매우 드문 상황이다. 그런 점에서, 중요한 근대이행기 텍스트 중의 하나인 대한매일신보의 국문논설을 대상으로 구축한 형태주석 말뭉치를 활용한 본고의 작업은 이미 성공적으로 독립신문 전산말뭉치를 구축, 활용한 연구의 후속작업으로서 그 의미가 있다.

또한 그동안 논설저자의 활동내용, 인식과 사상체계 및 문체 상의 질적 특성 등을 고려하는 식의 간접적인 방식으로 저자판별을 해 온 까닭에 논설저자와 관련된 논란이 계속되고 있는 현 상황에서, 말뭉치에 나타난 어휘들의 계량적 특성에 주목하여 대한매일신보 국문논설의 번역문체를 판별하는 어휘적 준거를 탐색하고자 하는 본 연구는 대한매일신보 논설저자들의 언론활동을 정확하게 이해하는 작업일 뿐만 아니라 이들의 언론에 관한 인식과 사상을 올바르게 파악할 수 있는 작업이 될 수 있다. 더 나아가 이들의 근대계몽기에서 차지하는 역사적 비중으로 보아 이들의 사회운동과 사상을 연구하는데 있어서도 정확한 저술을 대상으로 연구할 수 있는 객관적인 근거를 제시한다는 점에서 그 의미가 매우

큰 작업이 될 것이다.

## 2. 기존 연구 검토 및 말뚝치 구축방법

### 1) 대한매일신보 논설저자와 관련된 기존 연구 검토

지금까지 대한매일신보에 대해서는 창간과정과 운영(이광린, 1986; 정진석, 1987), 발행에 참여한 주요인물(정진석, 1987; 박정규, 2004, 김주현, 2006), 대한매일신보의 논설(유재천, 1986; 김덕모, 2004), 잡보(채백, 2004), 광고(안종묵, 2004), 독자(김영희, 2004) 등에 대한 연구가 수행되었다. 이러한 기존 연구에 의하면 대한매일신보의 논설은 총무로서 경영과 편집을 총괄했던 양기탁과 주필(지금의 논설위원)로 활동한 박은식, 신채호 등 당대의 대표적인 개신유학자들이 집필했던 것으로 알려져 있다(정진석, 1987; 박정규, 2004). 장도빈도 논설을 집필한 것으로 말해지고 있으나, 확실하지는 않다. 이 가운데 박은식은 1898년 9월 황성신문 창간 당시 주필이 되어 활동했고, 1904년 10월 17일 이전 대한매일신보 주필로 자리를 옮겨 1905년 8월 대한매일신보 국한문판이 발간되면서 주필로 활동하여 1907년 11월 5일까지 근무하다가, 다시 황성신문 주필이 되어 황성신문 폐간 당시까지 활동한 것으로 알려져 있다(김주현, 2008, 354~365쪽). 한편 신채호는 1906년 2월경부터 황성신문 주필로 활동하다가 1907년 11월 6일 대한매일신보 주필이 되었다(김주현, 2008, 365~367쪽).

그런데 지금까지 이들 대한매일신보의 주필 가운데 누가 어떤 논설을 썼는지 정확하게 밝혀지지 않은 실정이다. 이에 따라 대한매일신보 논설저자의 경우도 독립신문과 마찬가지로 대부분 간접적인 저자 판별 방식으로 논설저자를 판별해왔다. 이들의 활동조직, 활동내용, 인식과 사상체계 및 문체상의 특성 등을 고려하는 간접적인 방식으로 대한매일신보의 일부 무기명 논설이 신채호, 박은식 또는 양기탁의 글로 판정되

어 각각 전집에 수록된 것이다.

그러나 대한제국기 박은식 저작목록을 검토한 노관범(2002)은 1975년 간행된 『박은식전서』의 저작목록이 어떤 근거에서 작성했는지 합리적인 기준이 제시되지 않았고, 2002년 크게 증보된 『백암박은식전집』의 경우도 앞서 발행된 전서의 체계를 거의 그대로 수용하고 있다면서 황성신문과 대한매일신보의 무기명 논설 일부는 박은식의 저술로 추정할 수 있는 충분한 근거 없이 수록되었다고 지적했다. 그는 이어 박은식의 많은 저작들이 신문의 무기명 작품으로 묻혀 있을 것으로 예상된다면서 “이들 무기명 작품들을 박은식의 저작으로 판별할 수 있는 방법론의 모색”이 필요하다고 보았다(노관범, 2002, 287쪽). 신채호의 저작 여부를 검토한 김주현 역시 일부 무기명 논설이 “신채호의 문체나 사상과 유사하다고 여겨 전집에 포함시킨 것으로 보인다. 그러나 비슷하다고 하여 작품을 근거 없이 포함시키는 것은 잘못이다.”(김주현, 2006, 142쪽)고 지적했다. 실제 1977년에 간행된 기존 전집을 대폭 증보한 『단재 신채호전집』의 제6권(2008) 논설·사론의 해제는 기존 전집에 수록되어 있는 논설 및 사론에 문제가 심각하다고 하면서도 제6권에 수록한 무기명 신문논설의 신채호저작 판정의 객관적 기준을 제시하지는 않았다(김삼웅, 2008). 이런 문제들을 해결하기 위해 박정규는 저자판별을 위해 보다 체계적이고 과학적인 분석이 필요하다고 지적한 바 있다(박정규, 2004).

한편 대한매일신보의 주필이었던 박은식과 신채호가 많은 논설과 역사서들을 남겼는데 비해, 대한매일신보의 총무였던 양기탁은 저서가 남아 있지 않고 기명으로 된 논문도 거의 없다. 이에 대해 정진석(1997)은 그가 박은식, 신채호와 근본적으로 같은 사상을 가졌지만 이들과 달리 좀 더 현실적인 감각을 지닌 사람이었던 것이 원인이었다는 것으로 보았다. 그 원인이 무엇이든 실제 『우강 양기탁전집』(우강양기탁선생전집편찬위원회, 2002)에 수록되어 있는 신문논설과 잡지기사는 여섯 편에 불과하다. 이 중에서 잡지기사 두 편을 제외한 네 편의 신문사실도 두 편의 국한문논설(「百梅特捏이 不足以壓一伊太利」과 「學界의 花」)과

이를 번역한 국문논설이다. 그러나 이 두편의 경우도 그의 글이 아닐 가능성이 제기되고 있다(김주현, 2006, 126쪽). 그의 이름으로 쓴 글은 1906년 8월 『가명잡지』에 게재된 「가명교육론」이 유일하다. 이렇듯 양기탁의 경우엔 그가 남겼다고 전해지는 논설의 수가 매우 적기 때문에 인지, 양기탁의 원전에 대한 논란은 거의 찾아볼 수 없는 실정이다. 이는 그를 사상가로서가 아닌 독립운동가로서 인식했던 역사학계의 태도가 반영된 결과라고도 볼 수 있을 것이다.

장도빈에 대한 언론사학자의 평가는 대체로 회의적이다. 우선 정진석(1997, 137쪽)은 그를 신보사 기자로 출발하여 일제시대에도 언론계에 종사했던 인물 가운데 한 사람으로 설명하고 있는데, 그것은 김중희가 펴낸 『汕耘 장도빈』(1985)이 적지 않은 오류가 있어 활용에 유의할 필요가 있다고 보기 때문이었다. 이에 반해, 김창수(1988a)는 『思想界』에 발표된 장도빈의 회고록적 자전 “암운길은 구한말”(1962)과 『장도빈 전집』(1981, 1982)에 실린 천관우의 간행사를 근거로 하여, 장도빈이 대한매일신보의 주필로 활동한 것으로 설명했다. 그에 의하면 1888년생인 장도빈이 처음 언론인으로 입신하게 된 것은 당시 황성신문 주필 박은식의 소개로 대한매일신보 총무로 있던 양기탁에게 발탁되면서부터였다. 그리고 그가 신문사에 입사한지 얼마 되지 않아 1906년 이래 신문의 주필이며 장도빈 보다 8년 연상인 신채호가 와병 중이었기 때문에, 장도빈이 주필이 되어 단독으로 논설을 집필하기도 했고, 또 그 이듬해인 1909년부터는 신채호와 더불어 1주일씩 교대로 논설을 집필했다는 것이다(52쪽). 이러한 평가를 근거로, 그는 대한매일신보 논설에 실린 장도빈의 역사논문에 대한 검출작업이 요청된다고 지적하였다(53쪽). 또 김창수는 다른 글(1988b)에서 1910년 이후 신채호가 망명하면서부터 이 신문이 폐간될 때까지의 약 5~6개월간 장도빈이 단독으로 집필한 것으로까지 판단하여, 장도빈이 신채호의 와병으로 논설을 집필하기 시작하면서 집필했을 것으로 보이는 논설들을 대상으로 장도빈의 민족의식을 고찰하였다. 그런데 김창수는 장도빈의 문체와 관련해, 생전에 장도빈과

친교가 두터웠던 오천석이 장도빈과 신채호의 문체 특징을 비교해 장도빈의 문체에서는 ‘하얏다’, ‘오호(嗚呼)’ 등이 반드시 들어가는 경향이 있고 신채호의 경우엔 ‘하였다’고 표현하는 경향이 있음을 지적했다고 증언한 것은 본 연구와 관련해 주목할 만하다(75쪽). 그러나 그가 장도빈의 논설이라고 간주했던 논설 가운데 이미 양기탁의 논설로 평가받는 “학계의 화”를 포함시키고 있다는 점에서 그의 분석은 논설의 원전과 관련해 논란의 여지가 적지 않다. 이런 이유가 원인이 되었는지 장도빈의 논설들은 이미 발행된 『장도빈 전집』에 포함되어 있지 않을 뿐만 아니라, 장도빈의 언론활동에 대해서도 논란이 있는 것으로 이해된다.

또한 대한매일신보의 국문논설과 관련해서도 논란은 존재한다. 대한매일신보 국문논설은 주로 국한문판의 논설을 번역한 것으로 김연창(金演昶)이 번역을 담당한 것으로 알려져 있다(정진석, 1987, 137쪽). 그러나 국문판의 모든 논설이 국한문판 논설을 번역한 것은 아니다. 실제 국한문판에 게재되지 않은 논설로 국문판에만 게재된 논설들이 있다. 또한 김연창의 번역이 아닌 주필 자신이 한글로 집필한 논설일 가능성도 있다. 당시의 대표적인 주필이었던 신채호 관련해서만 살펴보더라도 기존 전집을 대폭 확충해 간행된 『단재신채호전집』(2008) 제6권 논설·사론편에는 신채호 집필여부 판정기준을 기명(記名), 인정, 추정 3종류로 구분하여 그가 집필한 것이거나 집필한 것으로 인정 또는 추정되는 논설 및 사론을 수록했다. 국한문판에 게재하기 전에 국문판에 먼저 게재된 논설도 2건이 있다. 또한 단재문화예술제추진위원회가 간행한 『단재신채호』(2006)에는 국한문판에는 게재되지 않고, 국문판에만 게재된 논설 가운데 신채호가 쓴 것으로 인정되는 논설들을 적지 않게 수록하였다.

이처럼 대한매일신보의 대표적인 논설필자인 박은식과 신채호의 경우는 물론 다른 관련인물들에 대해 현재까지 그들의 저술판정이 정확하게 되어 있지 않은 실정이다. 문제는 개화기 대표적인 언론인이면서 동시에 애국계몽운동기의 대표적인 사회운동가이며 사상가인 박은식이나 신채호, 양기탁에 관한 많은 연구들이 이렇게 편찬된 기존 전집들을

기본 문헌으로 하여 이루어지고 있다는 점이다.

## 2) 형태주석 말뭉치의 구축과정

근대이행기를 비롯한 특정시기의 텍스트들, 특히 신문에 대한 계량적인 내용분석을 위해 컴퓨터를 활용하는 연구들이 매우 활발히 전개되어 왔던 서구와 달리, 한국의 경우 계량적인 내용분석을 위한 형태주석 말뭉치 구축과 활용은 2000년을 전후해서야 비로소 학계의 관심을 받기 시작하였다. 이와 관련해 고려대학교 민족문화연구원이 1998년부터 ‘21세기 세종계획’ 사업을 통해 이룩한 성과는 주목할 만하다. 민족문화연구원은 이를 위해 1998년 문자코드연구센터를 수립하여, 21세기 정보환경 속에서 근대이전 시기의 옛한글을 검색 가능하도록 하기 위한 기획으로서 ‘옛한글 정보 데이터베이스 구축’에 착수하였다. 이를 위해 15세기부터 1920년대까지의 옛한글의 자형을 조사하는 한편, 시대별, 문헌성격 유형별, 장르별, 각 시대 문헌 성격 유형별, 각 시대 장르별 옛한글 출현빈도의 상관관계를 검증하였다. 그리고 그 결과를 『컴퓨터와 인문학 총서』로 간행하여, ‘21세기 세종계획’의 제2차년도 사업을 통해 구축된 150만 어절의 현대 국어 말뭉치를 기반으로 하여 국어 정보화의 기반이 되는 한국어의 형태소 및 어휘의 사용빈도에 관한 다양한 통계자료와 분석을 담은 『한국어 형태소 및 어휘 사용빈도의 분석 1』(김홍규·강범모, 2004)과 『한국어 형태소 및 어휘 사용빈도의 분석 2』(김홍규·강범모, 2004)를 간행하였다. 그러나 고려대 민족문화연구원 세종 역사자료 말뭉치 분과에서 구축한 약 1,100만 어절을 비롯하여 그밖에 한국학 중앙연구원의 ‘장서각 한글필사본’ 자료 310만 어절, 선문대 박재원 교수가 입력한 자료 등을 합치면 약 2,000만 어절이 될 것이라고 추정되고 있지만, 세종계획 말뭉치 외의 자료는 대부분 미공개 상태라서 다른 연구자들이 사용할 수 없을 뿐만 아니라 가공(加工)되지 않은 원시말뭉치이고, 또 세종계획 말뭉치의 규모 또한 당초 목표와 달리 약 63만 어절만이 구축되었다고 한다(김진해·차재은·김건희·이의철, 2009).



2000년을 전후로 한 시점부터 지금까지 축적되어 왔던 근대이행기 옛한글의 원시 말뭉치 및 형태주석 말뭉치는 국어학적·국문학적 의의와 활용을 목적으로 하여 대부분 소설 등과 같은 장르를 중심으로 한 구축이었다. 따라서 구한말 근대적 저널리즘의 장르까지는 포괄하고 있지 못하다는 점에서 직접적으로 구한말 신문의 내용을 계량적으로 분석하거나 검색하는 데 있어 적지 않은 한계점이 있다. 실제로 신문 장르에 대해 형태주석 말뭉치를 구축하고 이를 체계적으로 연구에 활용한 기존연구는 강남준 연구팀에 의한 성과와 연세대 홍윤표 교수에 의한 성과를 제외하곤 거의 전무한 실정이다. 강남준 연구팀은 2006년부터 연구를 개시한 이후, 오랜 기간에 걸쳐 독립신문의 논설에 대해 형태주석 말뭉치 구축작업을 수행했다(강남준 외, 2010). 또한 연세대 홍윤표 교수는 수년간의 수고를 통해 독립신문의 전체기사에 대한 형태주석 말뭉치를 구축하였다. 한편, 본 연구자들은 2010년 6월부터 11월에 이르는 상대적으로 짧은 기간 동안, 대한매일신보의 국문논설에 대한 형태주석 말뭉치를 구축하였다.

개화기 신문인 『대한매일신보』 국문논설의 정량적인 분석을 위해 선 분석 대상 데이터의 전산화 작업이 전제되어야 한다. 본 연구에서는 이미 구축된 국어사 자료의 형태 주석 말뭉치를 활용하여 『대한매일신보』 국문논설의 형태주석 작업을 진행하였다. 본 연구자들이 구축한 대한매일신보의 말뭉치는 1904년 8월 4일부터 1910년 7월 31일까지 총 877개의 국문논설을 대상으로 하였다. 산술적으로 6년에 해당되는 논설 수는 약 2,100여개이어야 하지만, 대상 논설수가 877개였던 것은 휴간기간이 있었을 뿐만 아니라 때에 따라 논설이 게재되지 않은 경우가 있었기 때문이다. 전체적인 형태주석 말뭉치의 구축과정은 다음과 같다<sup>1)</sup>.

말뭉치는 원문의 가공 여부에 따라 ‘원시 말뭉치(raw corpus)’와 ‘주석 말뭉치(annotated corpus)’로 구분한다. 따라서 말뭉치의 구축과정

1) 이하의 내용은 김영희·김민환·최운호·윤상길(2010)의 연구보고서 3장을 요약한 것이다.

은 크게 전산 원시말뭉치를 구축하는 첫 번째 과정과 이를 기반으로 형태주석 말뭉치를 구축하는 두 번째 과정으로 나뉜다.

우선 원시 말뭉치는 텍스트 자료를 원문의 특성을 그대로 보존하면서 구축되는 ‘기계 가독형 자료(machine-readable material)’로서, 여러 주석 작업의 기초 자료로 사용된다. 기계 가독형 자료가 되기 위해선, 가장 필수적인 입력 정보로 원문이 게재된 날짜, 기사 구분, 제목 등의 정보를 별도의 구획으로 입력해야 한다. 입력은 한글과 컴퓨터의 한글 워드 프로세서에서 유니코드 사용자영역에 배정하여 사용하고 있는 한글 옛글자 영역을 사용하였으며, 원시 자료의 입력에서 정보 표시 기호로 d(간행날짜), t(논설), ta(논설제목), type(띄어쓰기여부), p(단락구분)를 사용하였고, 정보 표시 기호와 해당 항목의 정보를 구분하는 구분자는 스트링 ‘##!’를 사용하였다. 다음은 1907년 11월 13일 『대한매일신보』에 게재된 “합심훈후 단테(합심한 후 단테)”라는 논설의 원시말뭉치의 일부를 예로 제시한 것이다.

```
d##!19071113
t##!론설
ta##!합심훈후단테
type##!spacing
p##!한국이 조리로 쇠약한식닭은 어디잇는뇨하면 단테가 못되는연고 | 오 지금까지 무익한
식닭이 어디잇는뇨하여도 단테가 못된연고 | 라
```

이러한 원시 말뭉치의 저장은 XML (eXtensible Markup Language) 문서로 최종적으로 저장하는 것이 일반적인데, 원시 말뭉치의 보관을 위해 텍스트 파일 형태의 XML 문서를 사용하는 이유는 XML가 문서의 구조를 사용자가 원하는 대로 정의하여서 사용할 수 있는 확장성(extensibility)을 제공해주고 말뭉치의 구조적 저장과 문서의 표현을 분리해주는 장점 때문이다.

이렇게 원시 자료로 입력한 텍스트와 이 자료를 가공해서 구조화한

구분	날짜/정보	문단	어절번호	어절
d	19071114			
t	론설			
ta	19071114	0	1	특군의
ta	19071114	0	2	확장
type	spacing			
p	19071114	1	1	청국은
p	19071114	1	2	본리
p	19071114	1	3	무비가
p	19071114	1	4	잇는
p	19071114	1	5	큰
p	19071114	1	6	나라으로
p	19071114	1	7	점점
p	19071114	1	8	변경이
p	19071114	1	9	되어

XML 문서는 모두 수평 말뭉치(horizontal corpus)이다. 한 행에 문장, 문단 등을 단위로 스트링이 수평으로 연결되어 있기 때문에 붙인 이름이다. 반면, 형태 주석을 위해서는 수직 말뭉치(vertical corpus)를 구성하는 것이 일반적이다. 다음은 수평말뭉치를 수직말뭉치로 구성한 예이다.

원시 말뭉치를 기반으로 형태주석 말뭉치를 구축하는 두 번째 과정은 용례에 기반하여 태깅을 붙이는 작업이다. 수직 말뭉치는 현행 한글 맞춤법을 기준으로 어절 단위 항목을 한 행에 배치해 놓았다. 각 행의 어절을 이미 분석해 놓은 용례가 있는 경우 해당 용례를 참조해서 자동으로 형태 주석을 첨가하고, 그렇지 않은 경우에는 참여 연구원들이 직접 형태 주석을 첨가한다. 형태 주석을 할 때 현대 한국어 맞춤법에 알맞게 표기한 현대어 표기로 번역한 자료를 함께 구축한다. 이때 현대 한국어로의 번역은 어휘 형태소(lexical morpheme)와 접두, 접미사만을 대상으로 한다.

다음 예는 형태 주석 말뭉치 중에서 1907년 11월 14일 논설만 추출한 것이다. 원시 어절에 대해서 형태 주석이 두 가지 항목으로 구성되어 있음을 알 수 있다. ‘특군의’라는 원시 어절은 어휘 형태가 현대어로 번역된 주석으로 ‘육군/NN+의/JK’로 분석되어 있으며, 원시 어절을 고어 분

d	19071114					
t	론설					
ta	19071114	0	1	특군의	육군/NN+의/JK	특군/NN+의/JK
ta	19071114	0	2	확장	확장/NN	확장/NN
type	spacing					
p	19071114	1	1	청국은	청국/NNP+은/JX	청국/NNP+은/JX
p	19071114	1	2	본리	본래/MA	본리/MA
p	19071114	1	3	무비가	무비/NN+가/JK	무비/NN+가/JK
p	19071114	1	4	잇논	잇/VA+는/ETM	잇/VA+는/ETM
p	19071114	1	5	큰	크/VA+ㄴ/ETM	크/VA+ㄴ/ETM
p	19071114	1	6	나라으로	나라/NN+으로/JK	나라/NN+으로/JK
p	19071114	1	7	점점	점점/MA	점점/MA
p	19071114	1	8	변경이	변경/NN+이/JK	변경/NN+이/JK
p	19071114	1	9	되어	되/VV+어/EC	되/VV+어/EC

석 방식으로 하여서 ‘육군/NN+의/JK’로 분석한 결과도 함께 구성하였다. 형태 분석 결과의 어휘 형태를 현대어로 표기한 대표형(citation form)으로 설정·번역하는 이유는 색인어를 추출하기 위해서이다.

### 3. 계량적 방법에 의한 저자판별과 어휘적 준거

#### 1) 계량적 방법에 의한 저자판별

계량적 방법에 의한 저자 판별연구(authorship study)는 텍스트를 작성한 저자들이 다른 저자들과 서로 다른 독특하고, 안정적인 특징을 가진 계량화할 수 있는 문체를 구사한다는 전제에서 출발한다(Holmes,

1985). 문체의 개념은 수사학, 문학, 언어학의 관점에 따라 변천을 겪었는데, 분야별로 서로 상이하게 정의되는 문체의 개념은 문체를 보는 관점의 차이에서 기인하였다. 문학을 전공하는 사람들은 문장으로부터 받는 다분히 주관적인 인상을 바탕으로 분류하거나 문장의 리듬 또는 이미지의 문제를 문체 연구의 대상으로 삼기도 했다. 그 후 문체의 연구는 현대 언어학의 발전으로 인해 다양한 방법론적 접근을 가능하게 했고, 글의 문체를 결정짓는 요인으로 순한글체와 국한문혼용체, 문어체와 구어체, 문체 인상(印象), 표기·번역 양식, 어휘와 구문, 글의 장르나 텍스트 유형, 사회적 요인, 작가와 작품 등 여러 가지가 지적되었다(장소원, 2005, 375쪽). 이렇게 문장 인상을 결정하는 데 참여하는 여러 가지 요인 중에서, 언어가 이용되는 방식이라고 할 수 있는 문체 요인은 저자 특성과 연결되어 있다.

문체 분석 통계기법의 기본은 언어 사용에 대한 확률적 분포에 대한 분석이다. 그리고 대상 말뭉치를 분석해 저자 판별을 하고자 할 때에는 어떠한 언어사용 단위가 가장 적절한 저자 판별의 문체소적 자질인가를 먼저 결정해야 한다(강남준 외, 2010, 83쪽). 이와 관련해 영어권의 계량적 문체분석에서는 일반적으로 저자 판별 목적으로 특정 주제와 관련된 내용어(content words)의 분포보다는 기능어(function words)의 분포가 더 중요한 판별기준이 된다고 지적되어 왔다(Mosteller & Wallace, 1984). 그러나 이러한 영어권의 논의는 대개 문법구분이 잘 발전된 영어 문장 분석에서나 적합하고 한국어 상황에서는 내용어와 기능어의 구분이 사실상 명확하기 않아 문제가 될 소지가 있지만, 신현숙(1999)은 한국어에도 기능어 형태가 있음을 지적하고 이를 네 가지 범주로 나누어 정의한 바 있다. ① 말이나 글 속에서 사람/사물/장소/사건/상태/방향 등을 가리키거나 꾸미는 기능을 하는 지시어 범주, ② 말이나 글 속에서 단어와 단어/문장과 문장/발화와 발화 등을 이어주거나 나열하는 기능을 하는 접속어 범주, ③ 통합하는 어휘 항목 특히 명사에 대한 통사 정보와 의미 정보, 나아가 언어 사용자의 인지 방법과 인지 태도를 포함하는 화용 정보까지 제공하는 조

사 범주, ④ 통합하는 어휘 항목 특히 동사에 대한 통사 정보와 의미 정보는 물론, 언어사용자의 인지 방법과 인지 태도를 포함하는 화용 정보까지 제공하는 어미 범주 등이 그것이다.

한편, 한나래(2009)의 연구는 『조선일보』칼럼의 저자판별을 위해 총 10가지 언어적 단위를 이용하여 저자판별을 시도하였다. 그 결과, 텍스트라는 가장 큰 문치에서부터 시작해서 단락, 문장, 구/절, 형태소, 음절, 그리고 마지막으로 음소에 이르는 다양한 단위 중에서, 형태소, 대표형 형태소, 비주제 특정 형태소(non-topic-specific morpheme)의 세 가지 형태소 기반 자질들이 저자 판별에 가장 적합하다는 결론에 이르게 되었다.

또한, 『독립신문』의 저자판별을 시도했던 강남준 외의 연구(2010)는 이러한 한나래의 연구가 주로 큰 틀에서 비슷한 주제를 다루고 있었던 사회·정치문제와 관련한 신문컬럼을 대상으로 한 탓에, 명사, 동사와 같이 내용어를 포함하는 형태소 혹은 대표형 형태소 자질을 이용할 수 있었다고 보았다. 따라서 백성들에 대한 교화, 신문물 소개와 같이 매우 다양한 주제를 다루고 있는 『독립신문』의 경우엔 이들 내용어가 저자의 언어습관이 아니라 글의 내용에 따라 저자 판별결과가 달라질 개연성이 있기 때문에, 조사의 사용 양상, 접속문을 구성하기 위한 연결어미, 관형형 어미 등이 이러한 문체적 특징을 나타내는 기능어가 될 수 있다고 지적하였다. 이에 따라, 선어말 어미를 준거로 주시경과 서재필의 저자판별을 시도하여 유의미한 연구결과를 도출하였다.

한편, 직접적으로 저자판별을 연구목적으로 삼지는 않았지만 형태주석 말뭉치를 문체연구에 도입한 연구로는 한영균(2009)의 연구가 있다. 한문과 한국어의 혼효어(混淆語)적인 요소의 극복은 궁극적으로 한국어의 문체가 그 현대성을 확보하는 과정에서 나타나는 어휘적 변화, 즉 각 텍스트에 쓰인 혼효적 요소들을 확인하고 그 변화양상을 살펴보는 것을 통해 이뤄질 수 있다고 본 그는, 1890년대 후반부터 1930년대 후반에 걸친 시기에 간행된 신문 기사 중 약 10년 주기로 각 시기별로 약

5,000어절 가량의 논설 기사를 추출하여, 국한문 논설과 국문논설의 각 시기별 어휘 변화 양상을 비교하였다. 특히 본고의 기획과 관련해 주목할 점은 국한문 혼용 문체의 현대성을 판별하기 위하여 사용한 어휘적 준거의 유형이다. 그는 문체 판별의 어휘적 준거의 유형으로서 총 아홉 가지를 제시하였는데, ① 대명사, ② 관형사, ③ 부사, ④ ‘단음절 한자+하’형 용언, ⑤ 단음절 한자어 체언, ⑥ ‘2음절 한자+하’형 용언, ⑦ 2음절 한자어 체언, ⑧ 한자어 감탄사, ⑨ 한자구가 그것이다.

지금까지 살펴본 기존의 저자 판별의 문체소적 자질에 대한 연구결과들을 종합해 볼 때, 한국어 상황에서 유의미한 문체소로 사용될 수 있는 것은 크게 두 가지이다. 첫째, 텍스트내용에 대한 통제가 이뤄질 경우 명사, 동사와 같이 내용어를 포함하는 형태소 혹은 대표형 형태소가 문체소적 자질을 가질 수 있다. 둘째 주제와 관련 없는 조사, 어미, 조동사, 의존명사, 형용사, 부사, 접속부사 등과 같은 비주제 형태소는 문체소적 자질이 될 수 있지만, 연구로서 그 실효성이 검증이 된 형태소는 선어말 어미일 뿐이며 나머지 형태소에 대해서는 검증이 필요하다.

통상적으로 문체 연구는 해당 저자의 계량적 ‘문체 특징(stylometric attribution)’을 우선 저자가 쓴 것이 확실한 문헌에서 추출한 다음 이것이 저자가 누군지 의심되는 문헌의 문체에서 나타난 특징과 비교해 저자 여부를 판정한다. 이때 사용되는 기법은 대부분 수리적 통계기법이며, 주성분분석(principal component analysis), 판별분석(discriminant analysis), 군집분석(cluster analysis) 등과 같은 다변량통계기법, 그리고 최근에는 연결망분석기법(network analysis)까지 다양하게 사용되고 있다(Grieve, 2007 ; 강남준 외, 2010, 82쪽에서 재인용). 그러나 본고가 분석하고자 하는 대한매일신보의 경우엔 논설의 원저자 뿐만 아니라 번역 저자에 대한 판별이 확정된 상황이 아니고, 또 기명논설 또한 제한적인 상황이다. 따라서 ‘일정한 기준’에 의거하여 각 문서 혹은 어휘들 간의 거리를 비교하는 군집분석<sup>2)</sup>을 통해 전체 논설을 일차적으로 분류하는 작업은 불가피하며, 그 ‘일정한 기준’ 마련의 차원에서도 문체소적 자질을

먼저 추출해 내는 작업은 필수적이다.

## 2) 분석방법 : 기대빈도와 관찰빈도 차이에 의한 어휘적 준거의 도출

앞의 기존 연구를 토대로 하여, 대한매일신보 국문논설의 번역저자를 판별하기 위한 탐색적 작업으로서 본고는 기구축된 『대한매일신보』 국문논설의 말뭉치에서 사용된 어휘를 분석한다. 일반적으로 저자 개인 문체에서 비롯되는 언어사용빈도 특징을 파악하기 위한 방법 중, 빈도에 의한 기댓값 계산은 가장 간단하면서도 다른 복잡한 통계 기법의 적용 없이 언어 자료를 해석할 수 있다는 장점이 있다(신효필, 2005). 따라서 본고에서는 전체 코퍼스에서 사용한 어휘 형태의 빈도를 이용하였다.

우선 각 시기에 나타난 어휘들의 관찰빈도(O)와 기대빈도(E)의 차이를 비율(O/E)로 계산하여서 이를 각 시기별로 비교, 특징적인 형태들을 관찰하였다. 이때 각 형태소별로 높은 빈도를 보이는 어휘들이나 낮은 빈도를 보이는 어휘들에 특별히 주목하였다. 고빈도 어휘들에 대한 분석을 통해서는 전체적인 경향성을 확인하고자 하였고, 각 시기별로 특별히 많이 사용된 어휘들에는 어떠한 것이 있었는지를 살펴보고자 하였다.

이러한 시기 간 어휘빈도 비교를 위해 본고는 시기를 네 시기, 즉 P00, P01, P02, P03으로 구분하였다. 이때 본고의 시기구분은 1) 대한매일신보의 ‘언어판’ 변화, 2) 논설 집필진의 변화, 3) 국문논설이 he 언어판 논설과 맺는 관계를 고려한 것이었다. 첫 번째 시기는 1904년 8월 4일부터 1905년 3월 9일까지의 시기이고, 두 번째 시기는 1907년 5월 23일부터 1907년 11월 5일까지의 시기, 세 번째 시기는 1907년 11월 6일부터 1908년 5월 13일까지의 시기, 네 번째 시기는 1908년부터 1910년 7월 31일까지의 시기이다.

---

2) 문서 혹은 어휘의 거리비교에 사용되는 대표적인 군집분석 방법으로는 ‘최소편집거리(minimum edit distance)’라고도 불리는 레벤스타인 거리(Levenshtein Distance) 등이 있다(최운호, 2010 ; 최운호 · 김동건, 2009).



첫 번째 시기는 대한매일신보의 ‘언어판’ 변화에 근거한 것이다. 이 시기에 대한매일신보는 영문판 4면, 한글판 2면을 합쳐 6면을 함께 발행하였다. 여기서 한글판의 사설은 모두 영문판 사설을 번역한 것이었던 것으로 이해될 수 있는데, 유재천(1990)의 「부록-1」에 제시되어 있는 두 언어판 논설 제목목록을 통해서도 이를 확인할 수 있고, 박정규(2004)의 연구에서도 당시 한국어를 잘 모르는 배설이 영문판에 기사와 논설을 집필하고 영어와 일어에 능통한 양기탁이 이를 번역한 것으로 보인다고 지적하고 있다.

두 번째 시기는 (자진휴간 이후 영문판과 국한문판을 분리·발행하던 시기를 거쳐) 국한문·한글·영문판 3종의 발행을 시작했던 시기 중 박은식이 논설주필로 활동하였던 것으로 알려진 시기이며, 세 번째 시기는 신채호가 논설주필로 활동하기 시작했다고 알려진 시기이다. 두 번째 시기와 세 번째 시기의 경우, 국문논설은 국한문논설을 김연창이 번역한 것으로 알려져 있지만(정진석, 1987, 137쪽), 모든 사설을 김연창이 번역하지 않았을 가능성이 있다(김영희 외, 2010, 53-54쪽).

네 번째 시기는 국한문판 논설과 국문판 논설이 동일 날짜에 게재되기 시작한 시기로서, 동 시기를 독자적으로 구분한 이유는 번역된 국문논설의 경우 하루 이틀 정도의 시간적 차이를 두었던 것으로부터 동일 날짜 게재로 바뀌었다는 사실 그 자체가 국문논설을 번역한 제3의 인물이 존재하였거나 또는 신채호가 국문논설을 직접 집필하였을 가능성이 있고, 또 그로부터 사용되는 어휘에 변화가 야기되었을 가능성이 있기 때문이다.

## 4. 분석결과

### 1) 고빈도 어휘의 특징

본 절에서는 대한매일신보 국문논설에서 사용된 어휘의 시기별 출현빈도 간의 차이가 유의미한 차이인지를 검토하기 위한 통계수치로서 각 시

기 어휘들의 기대빈도(E)에 대한 관찰빈도(O)의 비율(O/E) 수치를 활용하여, 전체 논설코퍼스에서 가장 많이 사용된 (고빈도)부사 어휘와 형용사 어휘의 특성을 제시하도록 하겠다. 그러나 여기서는 이미 지적인 바 있듯이 내용어(context word)는 주제에 따른 편차가 기능어(function word)에 비해 더 큰 특징을 가지고 있고, 특히 명사는 논설주제에 따라 많은 편차를 보이고 있었기 때문에 이에 대한 제시는 생략하기로 하겠다.

#### (1) 고빈도 어휘의 시기별 분포

〈표 1〉은 전체 논설 코퍼스에서 사용빈도가 높은 부사들이 각 시기별로 관찰빈도와 기대빈도의 비율이 어떻게 되는지를 보여주고 있다.

**표 1. 고빈도 부사의 각 시기 간 O/E 수치 비교**

morph_all	P00	P01	P02	P03
또	0.8460	0.8658	0.8001	1.0760
혹	0.5392	0.8543	1.1333	1.0857
다만	1.1076	0.4550	0.9885	1.0299
고로	0.8109	1.2159	0.6669	1.0726
다	1.2608	1.7784	1.1925	0.8444
이제	0.7077	1.2543	0.4386	1.1270
곧	0.4514	0.5157	1.1384	1.1338
비록	0.5733	0.7871	1.2922	1.0592
지금	1.7605	0.6528	1.5978	0.7802
서로	1.1191	0.7781	0.7939	1.0294
더욱	0.8013	1.0836	0.6142	1.0952
그러나	0.2900	0.5314	1.0327	1.1824
능히	0.7751	1.4947	0.9341	1.0112
<b>하어금</b>	<b>2.0094</b>	0.8704	0.6055	0.8699
과연	1.1616	0.7555	0.5409	1.0637
<b>필경</b>	<b>2.2679</b>	0.6055	0.9610	0.7840

다시	1,3380	1,1908	1,0123	0,9116
오히려	1,0328	0,3154	0,8427	1,0815
반드시	0,2320	0,9384	1,4360	1,0917
대저	0,1569	1,3961	1,0173	1,1328
<b>또한</b>	<b>4,5490</b>	0,1549	1,5805	0,2602
이미	0,0245	0,0545	0,2778	1,4022
점점	1,2972	0,7983	0,5221	1,0351
스스로	0,4450	1,1140	1,1926	1,0715
<b>심히</b>	<b>2,0826</b>	0,5009	0,3549	0,9293
더	1,8307	1,0991	1,0627	0,8115
실로	0,2847	0,4223	0,5983	1,2637
없이	2,0906	0,7502	0,4677	0,8866
많이	2,2021	1,5076	0,7690	0,7458
항상	1,4816	1,8130	0,6539	0,8833
먼저	0,8792	0,9356	0,5303	1,1064
함께	0,3195	0,2666	0,5036	1,2861
아무리	1,3480	0,5453	0,3091	1,0824
가장	0,1261	1,8712	1,4848	1,0200
깊이	0,8995	1,2640	1,3134	0,9457
오직	0,2859	0,8483	2,1034	0,9811
각각	0,8694	0,5373	1,2182	1,0337
하물며	0,0993	0,7732	0,5634	1,2749
홀로	0,1986	0,3314	0,7512	1,2648
감히	0,7880	0,6262	0,9937	1,0785
<b>아무쪼록</b>	<b>4,6435</b>	0,5438	0,3082	0,4110
<b>대단히</b>	<b>5,6435</b>	1,1161	0,8698	0,0639
그러므로	0,8372	0,8597	0,2436	1,1683
새로	0,9837	0,5837	0,9924	1,0427

위 표에서 볼드체로 표시한, ‘하여금’, ‘필경’, ‘또한’, ‘심히’, ‘아무쪼록’, ‘대단히’와 같은 고빈도 부사 어휘는 기대빈도보다 최소 약 2배에서 최대 약 5배 이상으로 많이 출현하였는데, 여기서 특징적인 것은 이러한 부사 어휘들이 모두 첫 번째 시기(P00)에 집중되어 있다는 점이다. 이러한 경향성은

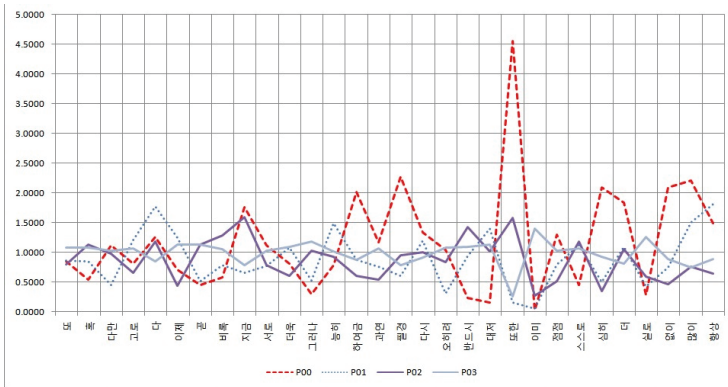


그림 1. 고빈도 부사의 각 시기 간 O/E 수치 비교 그래프

부사어 중에서 많이 사용된 상위 30개의 어휘들이 각 시기별 논설에서 어떻게 사용되었는지를 보여주고 있는 <그림 1>에서 직접적으로 확인된다.

그리고 첫 번째 시기에 특징적으로 많이 출현했던 6개의 고빈도 부사 어휘들을 각각 종류별<sup>3)</sup>로 구분해 보면, 문장(文章)부사가 3개(‘필경’, ‘또 한’, ‘아무쪼록’), 성분(成分)부사 중 성상(性狀)부사가 3개(‘하여금’, ‘심히’, ‘대단히’)가 있었다. 이 중 관심있게 보아야 할 어휘는 관찰빈도와 기대빈도 간의 비율(O/L)이 다른 부사와 비해 월등히 높은 ‘아무쪼록’과 ‘대단히’이다. ‘대단히’는 정도를 나타내는 정도부사이고, 아무쪼록은 “화자가 발화행위를 수행하는 방식을 나타내 주는”(김경훈, 1996, 43쪽) 수행부사이다.

한편, <표 2>는 전체 논설 코퍼스에서 사용빈도가 높은 형용사들이 각 시기별로 관찰빈도와 기대빈도의 비율이 어떻게 되는지를 보여주고

3) 부사의 종류는 수식하는 단위를 기준으로 크게 문장(文章)부사와 성분(成分)부사로 나뉘며, 또 성분부사는 다시 사람이나 사물의 모양, 상태, 성질을 한정하여 꾸미는 성상(性狀)부사와 용언의 앞에 놓여 그 내용을 부정하는 부정(否定)부사, 그리고 장소나 시간을 한정하여 나타내는 지시(指示)부사로 나뉜다. 한편, 부사는 수식대상과 맺는 관계의미론적 성격에 따라 정도부사, 상태부사, 평가부사, 양태부사 등으로 나뉘기도 한다(김경훈, 1996).

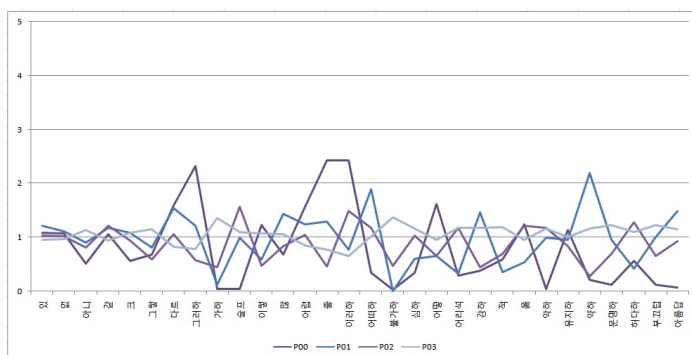
표 2. 고빈도 형용사의 각 시기 간 O/E 수치 비교

morph_all	P00	P01	P02	P03
있	1,0822	1,2092	1,0250	0,9600
없	1,0739	1,1041	1,0141	0,9731
아니	0,5087	0,8967	0,8043	1,1413
갈	1,0529	1,1760	1,2095	0,9392
크	0,5666	1,0738	0,9350	1,0921
그렇	0,6823	0,8096	0,5831	1,1496
다르	1,5897	1,5319	1,0598	0,8214
<b>그러하</b>	<b>2,3114</b>	1,2042	0,5687	0,7837
가하	0,0412	0,1221	0,4499	1,3648
슬프	0,0417	0,9902	1,5609	1,1056
이렇	1,2250	0,5865	0,4693	1,0778
많	0,6762	1,4329	0,8324	1,0534
어렵	1,5713	1,2294	1,0452	0,8552
<b>중</b>	<b>2,4189</b>	1,2853	0,4553	0,7727
<b>이러하</b>	<b>2,4213</b>	0,7696	1,4831	0,6533
어떠하	0,3499	1,8908	1,1662	1,0240
불가하	0,0337	0,0000	0,4677	1,3746
심하	0,3388	0,6030	1,0254	1,1670
어떻	1,6213	0,6559	0,6505	0,9614
어리석	0,2963	0,3296	1,1677	1,1777
강하	0,3864	1,4615	0,4386	1,1735
적	0,5892	0,3496	0,6934	1,1929
옳	1,2380	0,5331	1,2087	0,9605
악하	0,0404	0,9885	1,1714	1,1690
유지하	1,1362	0,9480	0,8358	1,0035
<b>악하</b>	0,2183	<b>2,1854</b>	0,2753	1,1679
문명하	0,1222	0,9516	0,6934	1,2329
허다하	0,5693	0,4223	1,2764	1,0960
부끄럽	0,1288	1,0030	0,6497	1,2339
아름답	0,0668	1,4865	0,9267	1,1575
높	0,2042	1,8177	1,2878	1,0408
부족하	1,1324	0,4724	0,8925	1,0387
희	0,3539	1,1023	1,4281	1,0531
용렬하	0,0722	0,4820	0,6374	1,2952

해롭	0,5777	0,4820	0,7285	1,1774
밝	0,1489	1,4911	1,1268	1,1080
공변되	0,0752	1,6742	1,5183	1,0430
비참하	0,1505	0,5023	1,5183	1,1350
<b>참혹하</b>	0,6844	<b>2,7072</b>	<b>1,6303</b>	0,8060
가석하	0,0769	0,1710	0,9693	1,2691
쉽	0,3108	1,2102	0,5879	1,1878
완전하	0,2383	2,1207	0,7011	1,1009
깊	0,0000	0,5361	2,0258	1,0805
유익하	1,6867	1,7871	0,8103	0,8186

있다. 표에서 볼드체로 표시한, ‘그러하’, ‘좋’, ‘이러하’, ‘악하’, ‘참혹하’와 같은 고빈도 형용사 어휘는 기대빈도보다 약 2배 정도 이상으로 많이 출현하였다. 앞서 살펴본 부사 어휘와 비교해 보았을 때, 첫 번째 시기(P00)에 집중되어 있는 형용사 어휘인 ‘그러하’, ‘좋’, ‘이러하’의 관찰빈도(O)와 기대빈도(E) 간의 비율(O/E) 수치가 현격히 작다는 점, 그리고 이들 형용사 종류가 흔히 많이 쓰는 지시형용사라는 점에서 시기 간 유의미한 차이가 없는 것으로 해석된다. 한편, ‘악하’, ‘참혹하’와 같은 평가형용사<sup>4)</sup>가 두 번째 시기(P01)에 기대빈도도 약 2배 정도 많이 출현하는 데, 이 또한 관찰빈도(O)와 기대빈도(E) 간의 비율(O/E) 수치가 현격히 작다는 점에서 시기 간 유의미한 차이가 없는 것으로 해석된다. 이러한 경향성은 형용사 중에서 많이 사용된 상위 30개의 어휘들이 각 시기별 논설에서 어떻게 사용되었는지를 그래프로 보여주고 있는 <그림 2>에서도 확인할 수 있다.

4) 한국어 형용사를 분류하는 방법은 그 기준에 따라 여러 가지이지만, 의미와 논항(論項)을 기준으로 분류하면 대체로 네 가지 종류가 있다(박병수, 2003, 128쪽). ‘있다/없다/많다/적다’ 등과 같이 존재유무를 의미하는 ‘존재형용사(existential adjectives)’, ‘아프다’ 등과 같이 감각상태를 의미하는 ‘감각형용사(sense adjectives)’, 주관적 태도를 의미하는 ‘평가형용사(evaluative adjectives)’, 사물의 성질을 서술하는 ‘서술형용사(property adjectives)’가 그것이다.



## (2) 연구결과의 해석

가장 많이 출현한 부사와 형용사 어휘들의 시기별 빈도 분포를 종합해 볼 때, 고빈도 부사의 어휘들은 주로 첫 번째 시기, 즉 대한매일신보가 영문판과 한글판을 합쳐 발행하였던 시기이자 배설이 집필한 영문판 논설을 양기탁이 번역하여 한글판에 실었던 시기에 특징적으로 많이 출현하였던 데 반해, 고빈도 형용사의 경우엔 몇몇 형용사가 특정 시기에 특징적으로 많이 출현하는 특징을 보여주고는 있으나 그 형용사가 흔히 많이 사용되는 형용사 종류에 해당되거나 그 시기별 차이도 그다지 크지 않다는 점에서 유의미한 차이라고 볼 수 없었다. 이와 같은 연구결과는 기존 언론사 연구의 결론과도 호응하는 측면이 많다. 즉, 고빈도 부사 어휘가 특징적으로 첫 번째 시기와 나머지 시기 사이에 차이가 나타났던 것은 각 시기 간 번역자의 차이(양기탁과 김연창)로부터 비롯되었다는 해석이 가능하다.

물론 특정한 부사어휘가 첫 번째 시기에만 특징적으로 많이 출현한 이유로서, 주제와 상관없는 비주제 형태소인 부사 중에서도 평가부사<sup>5)</sup> (각주 2번 참고)가 논설주제와 밀접한 관련성을 맺을 개연성이 전혀 없

표 3. 부사 '아무쪼록'이 출현했던 논설의 날짜와 제목

논설날짜	논설제목	논설날짜	논설제목
19040809	명예를 유지함이라	19041011	동경 안에 흠담 나는 일
19040812	국권의 위엄이라	<b>19041013</b>	<b>한성 도로 교량</b>
19040818	장삼씨의 문대 김론이라	<b>19041026</b>	<b>연초 전미권</b>
19040823	려순 항복에 거절이라	19041027	전쟁을 담척함이라
19040824	역부문제라	19041102	한국에 덕한 일본 정책
<b>19040901</b>	<b>참정 경질</b>	19041104	한국 형편과 일의 전정
19040902	한국에 일본 위력이라	19041109	고맙지 안케 하는 일
19040903	전호 련속	<b>19041114</b>	<b>한국의 니두 형편</b>
19040905	전호 련속	19041119	[아라사의 니지 동정] 전호 련속
19040906	전호 련속	19041122	[빨틱함디] 전호 련속
19040907	전호 련속	19041126	[일본의 반동력] 전호 련속
<b>19040908</b>	<b>역부 문제</b>	19041206	려순에 일본 통신
19040910	한국에 일인 형격	19041212	[비밀 사회] 전호 련속
19040912	가옥 건축	19041213	변동되는 풍설
19040917	중지 소문	<b>19041219</b>	<b>교육 학논 법</b>
19040929	일진회와 밋 장삼씨의 요구함이라	19041221	리응씨
19041006	일본의 합동하라 함이라		
19041010	아라사의 빨틱 함디	19070726	삼십년 간에 대한 변란 력스

\* 비고 : 이 표는 형태주석 말뭉치의 검색을 통해 작성된 것임.

는 것은 아니다. 그러나 앞의 연구결과에서 서술한 바처럼, 이 시기에 특징적으로 많이 출현했던 고빈도 부사에는 평가부사가 없다는 점, 그리고 화자가 발화행위를 수행하는 방식을 나타내 준다는 측면에서 다분히 문체소적 자질을 가진 것으로 추정할 수 있는 수행부사인 '아무쪼록'의 관찰빈도와 기대빈도 간의 비율(O/E)이 월등히 높았다는 점은 오히려 논설주제와의 관련성에 대한 해석이 설득력 없다는 사실을 반증해 준다.

5) 이러한 평가부사의 예로는 '다행히', '어처구니없게도' 등이 있다(김경훈, 1996, 35쪽).



실제로 ‘아무쪼록’이 출현한 논설들의 목록을 보여주는 <표 3>을 보면, 이 어휘가 논설주제와 상관없이 빈번히 출현되고 있다는 점을 확인할 수 있다. 언뜻 보면, 러일전쟁과 일본의 조선침탈 등과 같은 당시의 정치상황과 관련된 논설에서 많이 출현하는 것으로 이해할 수도 있지만, 볼드체로 표시한 논설은 내정(內政)과 관련된 논설이었다.

## 2) 시기별 어휘 특성에 대한 분석

고빈도 어휘를 중심으로 전체적인 경향성을 살펴보았던 앞 절과 달리, 여기서는 각 시기별로 특징적으로 많이 출현했던 부사와 형용사를 제시하도록 하겠다. (1)항에서는 각 시기에만 특징적으로 나타나는 부사와 형용사 목록만을 제시하고자 하며, (2)항에서는 (1)항에서 제시된 목록을 근거로 하여 특징적으로 많이 출현했던 어휘의 시기별 차이가 가지는 의미에 대해 논하고자 한다.

### (1) 각 시기별 부사 및 형용사 출현빈도

#### ① 영문판논설의 번역국문논설 시기

다음의 <표 4>는 기대빈도 이상으로 많이 사용된 부사들의 목록을 나타낸 표이다. 이 부사들은 다른 시기에는 전혀 혹은 거의 관찰되지 않는다는 점에서 특징적이다.

표 4. 부사 출현빈도의 각 시기 간 O/E 수치 비교

morph_all	P00	P01	P02	P03
이다지	7,1485	0,0000	0,0000	0,0000
우금껏	7,1485	0,0000	0,0000	0,0000
자못	7,1485	0,0000	0,0000	0,0000
조용히	7,1485	0,0000	0,0000	0,0000
차차	7,1485	0,0000	0,0000	0,0000
별로	7,1485	0,0000	0,0000	0,0000
분명코	7,1485	0,0000	0,0000	0,0000
관곡히	7,1485	0,0000	0,0000	0,0000
급조히	7,1485	0,0000	0,0000	0,0000
긴급히	7,1485	0,0000	0,0000	0,0000
저저이	7,1485	0,0000	0,0000	0,0000
힘껏	7,1485	0,0000	0,0000	0,0000
감사히	7,1485	0,0000	0,0000	0,0000
거저	7,1485	0,0000	0,0000	0,0000
그나마	7,1485	0,0000	0,0000	0,0000
기위	7,1485	0,0000	0,0000	0,0000
기이히	7,1485	0,0000	0,0000	0,0000
능멸히	7,1485	0,0000	0,0000	0,0000
도저히	7,1485	0,0000	0,0000	0,0000
맹렬히	7,1485	0,0000	0,0000	0,0000
무례히	7,1485	0,0000	0,0000	0,0000
무심코	7,1485	0,0000	0,0000	0,0000
번거히	7,1485	0,0000	0,0000	0,0000
세부득이	7,1485	0,0000	0,0000	0,0000
안녕히	7,1485	0,0000	0,0000	0,0000
오죽	7,1485	0,0000	0,0000	0,0000
일껏	7,1485	0,0000	0,0000	0,0000
적당히	7,1485	0,0000	0,0000	0,0000
적연히	7,1485	0,0000	0,0000	0,0000
정녕코	7,1485	0,0000	0,0000	0,0000
정말로	7,1485	0,0000	0,0000	0,0000
지원히	7,1485	0,0000	0,0000	0,0000
지우금	7,1485	0,0000	0,0000	0,0000
처처이	7,1485	0,0000	0,0000	0,0000

친밀히	7,1485	0,0000	0,0000	0,0000
튼튼히	7,1485	0,0000	0,0000	0,0000
평탄히	7,1485	0,0000	0,0000	0,0000
한결같이	7,1485	0,0000	0,0000	0,0000
한량없이	7,1485	0,0000	0,0000	0,0000
도리어	6,8736	0,0000	0,0000	0,0560
정녕	6,7017	0,0000	0,5634	0,0000
매우	6,6611	0,5422	0,0000	0,0497
응당	6,6611	0,0000	0,2049	0,0662
방금	6,2549	0,0000	0,0000	0,1821
역시	6,2161	0,6915	0,0000	0,1267
또다시	5,9571	0,0000	0,0000	0,2428

표 5. 형용사 출현빈도의 각 시기 간 O/E 수치 비교

morph_all	P00	P01	P02	P03
진적하	7,1485	0,0000	0,0000	0,0000
불호하	7,1485	0,0000	0,0000	0,0000
과다하	7,1485	0,0000	0,0000	0,0000
미타하	7,1485	0,0000	0,0000	0,0000
불명하	7,1485	0,0000	0,0000	0,0000
흡족하	7,1485	0,0000	0,0000	0,0000
격원하	7,1485	0,0000	0,0000	0,0000
분망하	7,1485	0,0000	0,0000	0,0000
불건하	7,1485	0,0000	0,0000	0,0000
실답	7,1485	0,0000	0,0000	0,0000
응속하	7,1485	0,0000	0,0000	0,0000
진선진미하	7,1485	0,0000	0,0000	0,0000
강렬하	7,1485	0,0000	0,0000	0,0000
고맙	7,1485	0,0000	0,0000	0,0000
관곡하	7,1485	0,0000	0,0000	0,0000
궁극하	7,1485	0,0000	0,0000	0,0000
궁금하	7,1485	0,0000	0,0000	0,0000

그만하-, 근간하-, 기막하-, 난하-, 느리-, 망패하-, 명민하-, 모질-, 무르-, 무애하-, 무용하-, 미명하-, 미쁘-, 미성년하-, 배송하-, 번거롭-, 불길하-, 불분명하-, 빈핍하-, 삼엄하-, 성실하-, 소홀하-, 속되-, 수상하-, 순편하-, 습하-, 심란하-, 안되-, 안온하-, 암전하-, 연구하-, 유족하-, 일정하-, 적합하-, 정묘하-, 지각없-, 진중하-, 촉급하-, 친절하-, 태만하-, 특출하-, 하여하-, 한량없-, 험준하-, 화호하-, 훌륭하-, 희활하-

한편, <표 5>는 P00 시기에만 나타나는 형용사 목록을 나타낸 표인데, 관찰빈도와 기대빈도의 비율이 7.1485이다. 즉 원래 예측했던 것보다 7배 이상 많이 P00 시기에 출현한 형용사 목록이다. 한편, 표 아래의 목록은 모두 관찰빈도와 기대빈도의 비율이 7.1485를 나타낸 형용사 목록을 나타낸 것이다.

## ② 박은식 논설주필 시기

다음의 <표 6>은 P01 시기에 특징적으로 사용되는 부사 목록이다. 또한 <표 7>은 P01 시기에 특징적으로 사용되는 형용사 목록이다.

**표 6. 부사 출현빈도의 각 시기 간 O/E 수치 비교**

morph_all	P00	P01	P02	P03
가만	0.0000	15.9051	0.0000	0.0000
공변되이	0.0000	15.9051	0.0000	0.0000
결결	0.0000	15.9051	0.0000	0.0000
누차	0.0000	15.9051	0.0000	0.0000
소불하	0.0000	15.9051	0.0000	0.0000
시방	0.0000	15.9051	0.0000	0.0000
아까	0.0000	15.9051	0.0000	0.0000
요란히	0.0000	15.9051	0.0000	0.0000
무직히	0.0000	10.6034	0.0000	0.4857
암만	0.0000	10.6034	0.0000	0.4857
귀중히	3.5742	7.9525	0.0000	0.0000

morph_all	P00	P01	P02	P03
사사로이	0,0000	7,9525	0,0000	0,7285
무궁히	0,0000	7,9525	0,0000	0,7285
적막히	0,0000	7,9525	0,0000	0,7285
굳이	0,0000	7,9525	4,5073	0,0000
독히	0,0000	7,9525	4,5073	0,0000
한날	0,6499	7,2296	0,0000	0,6623
이욕고	2,8594	6,3620	0,0000	0,2914
착설히	0,0000	6,3620	0,0000	0,8742
몹시	0,0000	6,3620	1,8029	0,5828
강박히	3,5742	5,3017	1,5024	0,0000
역력히	2,3828	5,3017	0,0000	0,4857
유유히	2,3828	5,3017	0,0000	0,4857
진작	2,3828	5,3017	0,0000	0,4857
구차히	2,3828	5,3017	3,0049	0,0000
친히	1,5886	5,3017	0,0000	0,6476
지금껏	1,1914	5,3017	4,5073	0,0000
괴이히	0,0000	5,3017	0,0000	0,9714
금축히	0,0000	5,3017	0,0000	0,9714
한층	0,0000	5,3017	0,0000	0,9714
간절히	0,6216	4,8407	0,7839	0,7602
단단히	1,4297	4,2414	1,2020	0,5828
아마	2,6807	3,9763	1,1268	0,3643
온전히	1,1914	3,9763	0,3756	0,7892
도시	0,0000	3,9763	0,0000	1,0928
미처	0,0000	3,9763	0,0000	1,0928
천천히	0,0000	3,9763	0,0000	1,0928
필연코	0,0000	3,9763	0,0000	1,0928
은연히	0,0000	3,9763	2,2537	0,7285
평안히	0,0000	3,9763	2,2537	0,7285
만홀히	0,0000	3,9763	6,7610	0,0000
너무	3,1771	3,5345	2,0033	0,1619
일절	2,3828	3,5345	0,0000	0,6476
몰수이	0,0000	3,5345	1,0016	0,9714
무고히	0,0000	3,5345	1,0016	0,9714

표 7. 형용사 출현빈도의 각 시기 간 O/E 수치 비교

morph_all	P00	P01	P02	P03
겁악하	0,0000	15,9051	0,0000	0,0000
구차하	0,0000	15,9051	0,0000	0,0000
위태롭	0,0000	15,9051	0,0000	0,0000
절원하	0,0000	15,9051	0,0000	0,0000
강개하	0,0000	15,9051	0,0000	0,0000
고단하	0,0000	15,9051	0,0000	0,0000
광패하	0,0000	15,9051	0,0000	0,0000
균박하	0,0000	15,9051	0,0000	0,0000
궁핍하	0,0000	15,9051	0,0000	0,0000
기겁	0,0000	15,9051	0,0000	0,0000
도량방자하	0,0000	15,9051	0,0000	0,0000
만만하	0,0000	15,9051	0,0000	0,0000
무강하	0,0000	15,9051	0,0000	0,0000
방해롭	0,0000	15,9051	0,0000	0,0000
분통하	0,0000	15,9051	0,0000	0,0000
여의하	0,0000	15,9051	0,0000	0,0000
자악하	0,0000	15,9051	0,0000	0,0000
정성하	0,0000	15,9051	0,0000	0,0000
죄송하	0,0000	15,9051	0,0000	0,0000
준수하	0,0000	15,9051	0,0000	0,0000
편소하	0,0000	15,9051	0,0000	0,0000
현연하	0,0000	15,9051	0,0000	0,0000
호기롭	0,0000	15,9051	0,0000	0,0000
순후하	0,0000	10,6034	0,0000	0,485691
유리하	0,0000	10,6034	3,0049	0,0000
공정하	0,0000	9,543	0,9015	0,437122
친하	0,0000	9,0886	1,2878	0,416307
지극하	0,0000	8,8361	0,0000	0,647588

### ③ 신채호 논설주필 시기

다음의 <표 8>은 P02 시기에 관찰빈도가 기대빈도보다 월등히 많

표 8. 부사 출현빈도의 각 시기 간 O/E 수치 비교

morph_all	P00	P01	P02	P03
연죽	0,0000	0,0000	9,0147	0,0000
각별히	0,0000	0,0000	9,0147	0,0000
괴히	0,0000	0,0000	9,0147	0,0000
글리	0,0000	0,0000	9,0147	0,0000
기꺼이	0,0000	0,0000	9,0147	0,0000
땅땅	0,0000	0,0000	9,0147	0,0000
맹세코	0,0000	0,0000	9,0147	0,0000
부질없이	0,0000	0,0000	9,0147	0,0000
서서히	0,0000	0,0000	9,0147	0,0000
써	0,0000	0,0000	9,0147	0,0000
어여뻐	0,0000	0,0000	9,0147	0,0000
엄정히	0,0000	0,0000	9,0147	0,0000
연이나	0,0000	0,0000	9,0147	0,0000
정신없이	0,0000	0,0000	9,0147	0,0000
지극	0,0000	0,0000	9,0147	0,0000
처량히	0,0000	0,0000	9,0147	0,0000
황차	0,0000	0,0000	9,0147	0,0000
열심히	0,0000	1,9881	7,8878	0,0000
만홀히	0,0000	3,9763	6,7610	0,0000
현저히	0,0000	0,0000	6,7610	0,3643
촉은히	2,3828	0,0000	6,0098	0,0000
또는	0,0000	0,0000	6,0098	0,4857
쉬	0,0000	0,0000	6,0098	0,4857
이른바	0,0000	0,0000	6,0098	0,4857
적확히	0,0000	0,0000	6,0098	0,4857
그런고로	2,7332	0,0000	5,5679	0,0000
공손히	0,0000	0,0000	5,4088	0,5828
쾌쾌히	0,0000	0,0000	5,4088	0,5828
날날이	0,4766	2,1207	4,8078	0,3886
굳이	0,0000	7,9525	4,5073	0,0000
독히	0,0000	7,9525	4,5073	0,0000
지금껏	1,1914	5,3017	4,5073	0,0000
간혹	3,5742	0,0000	4,5073	0,0000
경중히	3,5742	0,0000	4,5073	0,0000

morph_all	P00	P01	P02	P03
왕왕	3,5742	0,0000	4,5073	0,0000
이런고로	3,5742	0,0000	4,5073	0,0000
중차	3,5742	0,0000	4,5073	0,0000
설령	2,3828	0,0000	4,5073	0,2428
잠잠히	0,0000	0,0000	4,5073	0,7285
궁극히	0,0000	0,0000	4,5073	0,7285
당돌히	0,0000	0,0000	4,5073	0,7285
모호히	0,0000	0,0000	4,5073	0,7285
사사로	0,0000	0,0000	4,5073	0,7285
설마	0,0000	0,0000	4,5073	0,7285
애중히	0,0000	0,0000	4,5073	0,7285

이 출현한 부사 목록이며, <표 9>는 P02 시기에 특징적으로 사용되는 형용사 목록이다

**표 9. 형용사 출현빈도의 각 시기 간 O/E 수치 비교**

morph_all	P00	P01	P02	P03
뜨겁	0,0000	0,0000	9,0147	0,0000
명철하	0,0000	0,0000	9,0147	0,0000
영민하	0,0000	0,0000	9,0147	0,0000
요사하	0,0000	0,0000	9,0147	0,0000
출중하	0,0000	0,0000	9,0147	0,0000
기결하	0,0000	0,0000	9,0147	0,0000
녹록하	0,0000	0,0000	9,0147	0,0000
부질없	0,0000	0,0000	9,0147	0,0000
양순하	0,0000	0,0000	9,0147	0,0000
가감하	0,0000	0,0000	9,0147	0,0000
가늘	0,0000	0,0000	9,0147	0,0000
값없	0,0000	0,0000	9,0147	0,0000



경박하	0,0000	0,0000	9,0147	0,0000
괴벽하	0,0000	0,0000	9,0147	0,0000
교양하	0,0000	0,0000	9,0147	0,0000
극중하	0,0000	0,0000	9,0147	0,0000
기특하	0,0000	0,0000	9,0147	0,0000
변화불측하	0,0000	0,0000	9,0147	0,0000
시	0,0000	0,0000	9,0147	0,0000
신묘하	0,0000	0,0000	9,0147	0,0000
아리따	0,0000	0,0000	9,0147	0,0000
안정하	0,0000	0,0000	9,0147	0,0000
여산하	0,0000	0,0000	9,0147	0,0000
연연하	0,0000	0,0000	9,0147	0,0000
염치없	0,0000	0,0000	9,0147	0,0000
영약하	0,0000	0,0000	9,0147	0,0000
온순하	0,0000	0,0000	9,0147	0,0000
우유하	0,0000	0,0000	9,0147	0,0000
잔잔하	0,0000	0,0000	9,0147	0,0000
절등하	0,0000	0,0000	9,0147	0,0000
정실하	0,0000	0,0000	9,0147	0,0000
졸하	0,0000	0,0000	9,0147	0,0000
진덕하	0,0000	0,0000	9,0147	0,0000
처연하	0,0000	0,0000	9,0147	0,0000
처참하	0,0000	0,0000	9,0147	0,0000
패약하	0,0000	0,0000	9,0147	0,0000
해연하	0,0000	0,0000	9,0147	0,0000
휘하	0,0000	0,0000	9,0147	0,0000
희호하	0,0000	0,0000	9,0147	0,0000
부당하	1,7871	0,0000	6,761	0,0000

#### ④ 동일 날짜에 국한문논설과 국문논설이 개제된 시기

다음의 <표 10>은 P03 시기에 관찰빈도가 기대빈도보다 높은 부사 목록이다. 이 시기에 나타나는 부사는 기대빈도와 거의 유사하게 관찰빈도가 나타나기 때문에 이 시기 논설문에 나타난 부사의 특징을 드러낸다

표 10. 부사 출현빈도의 각 시기 간 O/E 수치 비교

morph_all	P00	P01	P02	P03
대범	0.0000	0.0000	0.0000	1.4571
바야흐로	0.0000	0.0000	0.0000	1.4571
행여	0.0000	0.0000	0.0000	1.4571
분연히	0.0000	0.0000	0.0000	1.4571
익히	0.0000	0.0000	0.0000	1.4571
비상히	0.0000	0.0000	0.0000	1.4571
그저	0.0000	0.0000	0.0000	1.4571
거연	0.0000	0.0000	0.0000	1.4571
근근이	0.0000	0.0000	0.0000	1.4571
완연히	0.0000	0.0000	0.0000	1.4571
겹겹이	0.0000	0.0000	0.0000	1.4571
근근	0.0000	0.0000	0.0000	1.4571
만만	0.0000	0.0000	0.0000	1.4571
조금씩	0.0000	0.0000	0.0000	1.4571
가끔	0.0000	0.0000	0.0000	1.4571
덜	0.0000	0.0000	0.0000	1.4571
매매히	0.0000	0.0000	0.0000	1.4571
몸소	0.0000	0.0000	0.0000	1.4571
삼가	0.0000	0.0000	0.0000	1.4571
자자히	0.0000	0.0000	0.0000	1.4571

고 보기는 어렵다. 다만, 다른 시기에는 나타나지 않고 P03 시기에만 사용되는 부사가 있다는 것만 확인하는 수준이다.

한편 P03 시기의 형용사는 〈표 11〉에서 보듯이 그다지 특징적이지 않다. 즉, 원래 예상했던 기대 빈도에서 관찰 빈도가 많이 벗어나지는 않은 채 출현하고 있다.

## (2) 연구결과의 해석 및 과제

(1)항에서 제시한 부사와 형용사의 목록은 각 시기에만 특징적으로 관찰되는 어휘목록이었는데, 여기서는 이러한 연구결과에 대한 여러 가

표 11. 형용사 출현빈도의 각 시기 간 O/E 수치 비교

morph_all	P00	P01	P02	P03
간난하	0.0000	0.0000	0.0000	1.457073
우승하	0.0000	0.0000	0.0000	1.457073
외롭	0.0000	0.0000	0.0000	1.457073
광명하	0.0000	0.0000	0.0000	1.457073
푸르	0.0000	0.0000	0.0000	1.457073
혼미하	0.0000	0.0000	0.0000	1.457073
흑암하	0.0000	0.0000	0.0000	1.457073
찬란하	0.0000	0.0000	0.0000	1.457073
장려하	0.0000	0.0000	0.0000	1.457073
난만하	0.0000	0.0000	0.0000	1.457073
소소하	0.0000	0.0000	0.0000	1.457073
웅위하	0.0000	0.0000	0.0000	1.457073
맑	0.0000	0.0000	0.0000	1.457073
비창하	0.0000	0.0000	0.0000	1.457073
궁하	0.0000	0.0000	0.0000	1.457073
나약하	0.0000	0.0000	0.0000	1.457073
참독하	0.0000	0.0000	0.0000	1.457073
강건하	0.0000	0.0000	0.0000	1.457073
서늘하	0.0000	0.0000	0.0000	1.457073
찌르	0.0000	0.0000	0.0000	1.457073

지 가능한 해석을 제기하고 향후 과제를 논의하고자 한다.

우선 (1)항에서 제시한 두 번째 시기와 세 번째 시기의 특징적인 부사와 형용사 어휘에 대해서는, 크게 두 가지의 해석이 가능할 것으로 보인다. 첫 번째는 국한문논설의 국문번역자가 바뀐으로서 가능했다고 해석하는 것인데, 대한매일신보의 국문번역자가 김연창이라는 사실을 밝힌 정진석(1987)의 연구가 1908년 5월 28일자 시점의 일본공사관 기록에 근거하고 있다는 점에서 생각해 보면, 그 개연성은 충분해 보인다. 두 번째는 (전체 코퍼스를 대상으로 산출한) 기대빈도가 낮은 어휘가 상당수였던 것인 만큼 간헐적으로 제3의 인물이 번역했거나 직접 집필했을

가능성이다. 실제로 당시의 대표적인 주필이었던 신채호 관련해서만 살펴보다더라도 기존 전집을 대폭 확충해 간행된『단재신채호전집』(2008) 제6권 논설·사론편에는 신채호 집필여부 판정기준을 기명(記名), 인정, 추정 3종류로 구분하여 그가 집필한 것이거나 집필한 것으로 인정 또는 추정되는 논설 및 사론을 수록했다. 국한문판에 게재하기 전에 국문판에 먼저 게재된 논설도 2건이 있다.

한편, (1)항의 결과를 통해 도출할 수 있는 또 하나의 특징은 유독 네 번째 시기에만 특징적으로 사용되는 부사와 형용사가 거의 없었다는 점이다. 이에 대한 해석은 크게 두 방향에서 내려질 수 있다. 첫 번째는 이전시기의 국문번역자가 네 번째 시기까지 줄곧 번역을 담당했을 가능성이고, 둘째는 (담당자가 바뀌었다면) 국한문논설의 국문번역자가 점차 사용하는 부사와 형용사 어휘를 표준화해 나갔을 가능성이다.

이러한 해석가능성은 단지 가능성일 뿐이기 때문에, 이에 대해 설득력 있는 해명이 이뤄질 필요가 있다. 이런 측면에서 절실히 요구되는 작업은 국어학이나 언어학(방언학)의 견지에서 (1)항의 연구결과로서 제시되었던 어휘들에 대해 그 어휘가 일반적 표현에 사용되는 것인지 아니면 특수한 표현에 사용되는 것이지에 대한 실증적인 규명이 선행되어야 할 것으로 보여 진다. 또한 1)절에서 살펴 본 고빈도 어휘를 제외한다면, (본 절에서 살펴 본) 각 시기 특징적인 어휘들의 상당수는 관찰빈도가 상대적으로 낮았다는 점에서 이러한 결과가 통계 수치 상의 과장일 가능성이 오히려 높다는 해석도 가능하다.

## 5. 결론

계량적 방법에 의한 저자 판별연구에 있어, 언어 사용의 확률적 분포에 대한 분석은 기본적인데, 이 때 어떠한 언어사용 단위가 가장 적절한 저자판별의 문제적 자질인가를 결정해야 한다. 특히 논설의 원저자뿐만 아

나라 번역저자에 대한 판별이 확정된 상황이 아닌 대한매일신보의 계량적 문체 분석을 위해선 이 문체소적 자질의 추출과 이를 기준으로 한 군집분석은 불가피하다. 이에 따라, 본고는 대한매일신보 국문논설의 번역문체 판별의 어휘적 준거(문체소적 자질)를 탐색하기 위하여, 2010년 6월부터 5개월간에 걸쳐 구축하였던 대한매일신보 국문논설 형태주석 말뭉치를 대상으로 어휘들의 계량적 특성을 분석하였다.

분석결과, 대한매일신보가 영문판과 한글판을 합쳐 발행하였던 시기이자 배설이 집필한 영문판 논설을 양기탁이 번역하여 한글판에 실었던 시기에 6개의 고빈도 부사가 기대빈도에 비해 특징적으로 많이 관찰되었다. 특히 화자가 발화행위를 수행하는 방식을 나타내주는 수행부사인 ‘아무쪼록’의 관찰빈도와 기대빈도 간의 비율(O/E)은 월등히 높았다. 반면, 고빈도 형용사의 경우엔 몇몇 형용사가 특정 시기에 특징적으로 많이 출현하는 특징을 보여주고는 있으나 그 형용사가 흔히 많이 사용되는 형용사 종류에 해당되거나 그 시기별 차이도 그다지 크지 않다는 점에서 유의미한 차이라고 볼 수 없었다. 이러한 분석결과에 대해, 본고는 사용된 어휘의 시기별 차이가 번역자의 차이(양기탁과 김연창)로부터 비롯되었다고 해석하였다.

이러한 종합적인 분석결과로부터, 본고는 (대한매일신보 국문논설에만 국한시켜 적용시킨다는 전제 하에) 기능어 어휘 중에서 부사(副詞) 어휘가 문체적 특성을 어느 정도 나타내는 문체소적 자질이 될 수 있음을 확인하였다. 보다 구체적으로는 수행부사와 같은 특정 종류의 부사가 문체소적 자질이 될 수 있음을 보여 주었다.

그러나 이러한 본고의 연구결과는 영문판과 한글판을 합쳐 발행하였던 시기와 그 외 시기에 번역자가 달랐다는 기존 언론사 연구에 입각하여 첫 번째 시기에 대해서만 확인한 결과이며, 나머지 시기의 부사와 형용사 어휘에 대해서는 명확한 결론에 이르지 못하였다. 따라서 향후 대한매일신보 국문논설 형태주석 말뭉치를 본격적으로 활용하기 위해선, 국문학과 언어학 학계와의 협조 하에 대한매일신보 국문논설에 나타

난 여러 어휘와 개념들에 대한 이론적·실증적 해명이 담긴 향후 연구가 필요하다. 또한 저자 논란이 여전히 남아 있는 대한매일신보에 대한 계량적인 저자 판별연구를 본격적으로 수행하기 위해선, 본고의 연구결과를 활용한 군집분석이 실시될 필요가 있다. 이에 대한 작업은 추후의 과제로 남긴다.

## 참고문헌

- 강남준·최운호·이종영 (2010). 독립신문 논설의 형태 주석 말뭉치를 활용한 논설 저자 판별 연구: 어미 사용빈도 분석을 중심으로. 『한국사전학』, 15호, 73~101.
- 국립국어원 (2004). 『21세기 세종계획 국어기초자료 구축』.
- 김경훈 (1996). 『현대국어 부사어 연구』. 서울대학교 박사학위논문.
- 김덕모 (2004). 대한매일신보 논설 분석. 한국언론사연구회 엮음, 『대한매일신보 연구』 (185~265쪽). 서울: 커뮤니케이션북스.
- 김삼웅 (2008). 단재 신채호전집 제6권 논설·사론 해제. 단재신채호전집편찬위원회 편, 『단재신채호 전집』 제6권 (1~11쪽). 독립기념관 한국독립운동사연구소.
- 김영희 (2004). 대한매일신보 독자의 신문 인식과 신문접촉 양상. 한국언론사연구회 엮음, 『대한매일신보 연구』 (339~378쪽). 서울: 커뮤니케이션북스.
- 김영희·김민환·최운호·윤상길 (2010). 『대한매일신보 국문논설 형태주석 전산코퍼스의 구축과 활용』. 한국언론진흥재단 자유2010-01.
- 김진해·차재은·김건화·이의철 (2009). 역사자료 형태분석 프로그램 개발의 국어학적 의의와 활용 연구. 『어문연구』, 144호, 137~162.
- 김주현 (2006). 단재 신채호의 자료 발굴 및 원전확정 연구: 대한매일신보 소재 작품을 중심으로. 『한국현대문학연구』, 20집, 113~146.
- 김주현 (2008). 황성신문 논설과 단재 신채호. 『어문학』, 101집, 359~391.
- 김창수 (2008a). 汕耘 張道斌의 민족주의 사학( I ). 산운학술문화재단 엮음, 『산운 장도빈의 생애와 사상』 (49~70쪽). 서울: 산운학술문화재단.
- 김창수 (2008b). 汕耘 張道斌의 史學과 민족의식: 그의 논설에 나타난 歷史像. 산운학술문화재단 엮음, 『산운 장도빈의 생애와 사상』 (71~102쪽). 서울: 산운학술문화재단.
- 김홍규·강범모 (2004a). 『한국어 형태소 및 어휘 사용빈도의 분석 1』. 서울: 고려대 민족문화연구원.
- 김홍규·강범모 (2004b). 『한국어 형태소 및 어휘 사용빈도의 분석 2』. 서울: 고려

대 민족문화연구원.

노관범 (2002). 대한제국기 박은식 저작목록의 재검토 『한국문화』, 30, 263 ~307.

단국대학교 동양학연구소 편 (1975). 『박은식 전서』. 상-하. 서울: 단국대학교 동양학연구소.

단재문화예술제추진위원회 (2006). 『단재신채호』. 단재문화예술제추진위원회.

단재신채호선생기념사업회 (1977). 『단재 신채호 전집』. 상-하. 서울: 형설출판사.

단재신채호전집편찬위원회 편 (2007-2008). 『단재 신채호 전집』. 독립기념관 한국독립운동사연구소.

박병수 (2003). 몇 가지 종류의 한국어 형용사에 대하여. 한국언어정보학회 2003년도 학술대회 발표논문집. 128 ~131.

박정규 (2004). 대한매일신보의 참여인물과 언론활동. 한국언론사연구회 엮음, 『대한매일신보 연구』 (66 ~112쪽). 서울: 커뮤니케이션북스.

박정규 (2006). 국내에서의 신채호 연보와 쓴 글에 대한 고찰. 단재문화예술제추진위원회, 『단재신채호 연구의 재조명』 (59 ~74쪽). 단재문화예술제추진위원회.

백암박은식선생전집편찬위원회 편 (2002). 『백암 박은식전집』. 서울: 동방미디어.  
산운학술문화재단 (1988). 『산운 장도빈의 생애와 사상』. 서울: 산운학술문화재단.  
신현숙 (1999). 한국어 기능어의 어휘 정보 구축: 지시어. 『국제한국어교육학회』, 10권 2호, 215 ~232.

신효필 (2005). 언어자료의 통계분석과 관련된 몇 가지 고려사항들. 『어학연구』, 41호, 655 ~682.

안종묵 (2004). 대한매일신보의 광고에 관한 연구. 한국언론사연구회 엮음, 『대한매일신보 연구』 (291 ~338쪽). 서울: 커뮤니케이션북스.

우강양기탁선생전집편찬위원회 편 (2002). 『우강 양기탁전집』, 제1-4권. 서울: 동방미디어.

유재천 (1986). 대한매일신보의 논설분석. 이광란·유재천·김학동, 『대한매일신보 연구』 (51 ~137쪽). 서울: 서강대 인문과학연구소.

유재천 (1990). 『한국 언론과 이데올로기』. 서울: 문학과 지성사.

장소원 (2005). 현대국어와 역사성: 문체. 『국어학』, 45집, 373 ~418.



- 정진석 (1987). 『대한매일신보와 배설: 한국문제에 관한 영·일외교』. 서울: 나남.
- 정진석 (2004). 대한매일신보 창간의 역사적 의의와 그 계승문제. 한국언론사연구회 엮음, 『대한매일신보 연구』 (3~65쪽). 서울: 커뮤니케이션북스.
- 채 백 (2004). 대한매일신보 잡보의 내용분석. 한국언론사연구회 엮음 (2004), 『대한매일신보 연구』 (266~290쪽). 서울: 커뮤니케이션북스.
- 최운호 (2010). 레벤시타인 거리를 이용한 방언거리의 계산과 몽골어의 분석. 『알타이학보』, 20호, 127~160.
- 최운호·김동건 (2009). 군집분석 기법을 이용한 텍스트의 계통 분석: 수궁가 ‘고고천변’ 대목을 대상으로. 『인문논총』, 62집, 203~229.
- 한나래 (2009). 빈도 정보를 이용한 한국어 저자 판별. 『인지과학』, 20권 2호, 225~241.
- 한영균 (2009). 문체 현대성 판별의 어휘적 준거와 그 변화: 1990년대~1930년대 논설문의 한자어 사용양상을 중심으로. 『口訣研究』, 제23집, 305~342.
- Grieve, J. (2007). Quantitative authorship attribution: An evaluation of techniques. *Literary and Linguistic Computing*, 22(3), 251~270.
- Holmes, D., Robertson, M., & Paez, R. (2001). Stephen Crane and the New-York Tribune: A case study in traditional and non-traditional authorship attribution. *Computers and the Humanities*, 35(3), 315~331.
- Kenny, A. (1986). *A stylometric study of the New Testament*. London: Oxford University Press.
- Mosteller, F., & Wallace, D. (1984). *Applied bayesian and classical inference: The case of the Federalist papers*. NY: Springer Verlag.

최초 투고일 • 2011. 01. 09  
 논문 수정일 • 2011. 02. 08  
 게재 확정일 • 2011. 02. 10